

WEEKLY REPORT

张建伟

November 24, 2018

1 上周工作

1.1 三期 CT 肝肿瘤分割

- 把 DenseUNet 从 167 层减少到 125 层, 减轻模型过拟合, 基本保持了肝脏和肿瘤分割精度不变.
- 门静脉期单期分割结果: Liver: 0.936; Tumor: 0.609. 动脉期单期分割结果: Liver: 0.927, Tumor: 0.500
- 本周周一组会上讨论了一下, 在当前的数据量和 start-of-the-art 下, 用深度学习做全自动的肝肿瘤分割难以实现一个让医生可接受的分割效果.

2 下周工作

- 做文献检索 (1) 如何使用 (融合) 腹部的三期数据 (2) 肝肿瘤分割的半监督方法
- 完成秋学期数据挖掘课的课程论文

3 论文阅读

3.1 k-means++ The Advantages of Careful Seeding

本文针对传统的 k-means 算法中存在的问题, 提出了一些改进, 称为 k-means++ 算法. 传统的 k-means 算法在选取初始的聚类中心时是从数据集

中随机选择的数据点. 有文献提出这样的选择可能会导致极差的聚类结果. 所以 k-means++ 算法针对这个问题, 提出了一种新的选择初始聚类中心的方法, 其核心思想是初始聚类中心尽可能均匀的分布在整个数据集中. 方法如下:

- 1a. 从数据集 X 中按照均匀分布随机选择点作为中心 c_1 .
- 1b. 再选择一个中心 c_i , 其中数据点 $x \in X$ 满足分布 $D(x)^2 / \sum_{x \in X} D(x)^2$
- 1c. 重复步骤 1b, 直到选出了 k 个中心.

3.2 Towards an Optimal Subspace for K-Means

本文提出了子空间 k-means 的算法, 我们把特征空间分解为两个正交的子空间, 其中一个是聚类子空间, 另一个是噪声子空间. 然后把数据映射到聚类子空间后再实施 k-means 算法以排除噪声的干扰. 本文给出了子空间维度 m , 旋转矩阵 V 的一个简化的优化方法, 通过特征分解规避了大量的矩阵计算.

3.3 Discovering Non-Redundant K-means Clusterings in Optimal Subspaces

本文是上一篇论文的续作, 本文认为一个数据集可能有多种聚类子空间, 这些子空间相互独立, 从而也就会产生多种聚类方法. 本文的工作就是设计算法去发现多种聚类子空间. 其整体的算法思路和上一篇文章类似, 其中不同的地方在于由于多个子空间不能全部同时优化, 而他们两两正交, 所以在每步迭代时遍历每一种两个子空间的组合, 同时也把投影矩阵 P 加入了参数更新的过程.